APPLICATION OF STATISTICAL METHODS TO NAVAL OPERATIONAL TESTING

ALWYN SMITH, JR.

Mont 150

UNITED STATES
NAVAL POSTGRADUATE SCHOOL

APPLICATION OF STATISTICAL METHODS TO
NAVAL OPERATIONAL TESTING

by

Alwyn Smith, Jr.
Lieutenant Commander, United States Navy

APPLICATION OF STATISTICAL METHODS TO NAVAL OPERATIONAL TESTING

by

Alwyn Smith, Jr.

Lieutenant Commander, United States Navy

Submitted in partial fulfillment
of the requirements
for the degree of
MASTER OF SCIENCE

United States Naval Postgraduate School
Monterey, California

1956

APPLICATION OF STATISTICAL METHODS

TO

NAVAL OPERATIONAL TESTING

* * * *

Alwyn Smith, Jr.

This work is accepted as fulfilling

the thesis requirements for the degree of

MASTER OF SCIENCE

from the

United States Naval Postgraduate School

PREFACE

This thesis is concerned with the application of statistical methods to naval operational testing, and more specifically, with testing of the type conducted by OpDevFor and similar testing agencies. The statistical methods considered are those of confidence limits, sequential analysis and the more recently developed Statistical Decision Theory. These techniques are regarded as of interest to Naval line officers in various categories of billets, particularly Project Officers at testing agencies and officers concerned with planning which is based on the results of testing programs. Frequently, such officers are unfamiliar with the use and limitations of the three methods and the relations of these methods to each other.

Statistical Decision Theory, in particular, can be useful at higher levels of the naval establishment such as the offices of CNO. Naval planners at this level may be faced with a difficult problem in connection with testing programs whose objective is the estimation of the percentage effectiveness or probability of success of a weapon, in future combat. Such programs are costly to conduct, and increasingly so when an expensive weapon is tested to destruction. Costs of a different nature are those associated with the possible consequences if a poor estimate of the weapon's effectiveness is obtained from tests. The fundamental problem is to determine how many trials are to be conducted and hence how many weapons should be tested. Attempts to solve the problem by reconciling the conflicting costs will generally lead to a

dilemma. The application of Statistical Decision Theory to this problem is contingent upon the ability of the planner to specify the inputs or data required by the theory. An essential objective of this thesis is to show how a planner might be guided in specifying these inputs.

The thesis has been written with a view towards its usefulness for personnel with a minimum background in probability and statistics. It is addressed also to students and practicioners of Operations Analysis who may be concerned with the relations of the three statistical methods which are considered as tools which may provide quantitative basis for executive decision.

The writer's interest in the possible applications of this theory was aroused during the study of Statistical Decision Theory at the U. S. Naval Postgraduate School. The need for investigations as to how the inputs could be specified was pointed out by LT. R. A. Tucker, USN in his thesis: An Introduction to Statistical Decision Functions [1]. Tucker's paper presents detailed discussions of the mathematical concepts involved in the theory and the precise mathematical steps required to obtain the solution. It is intended for readers with less mathematical background than is required for an understanding of the basic work Statistical Decision Functions by Abraham Wald [2] . Readers who are interested in the theory and detailed computations should refer to Wald and Tucker.

This paper is divided into five chapters. Chapter I discusses the relationship between types of operational testing problems at various levels of the Navy. Chapters II and III are examples of applications

of statistical methods at the testing agency level. In Chapter IV, an example of a guided missile is used as a vehicle of discussion as to how the inputs required by Statistical Decision Theory may be specified by an office of CNO. A solution to the example is then given. In Chapter V the effect of variation of parameters is shown.

This thesis was written at the U. S Naval Postgraduate School, Monterey, California, during the period January-May, 1956. I am indebted to Professor Thomas E. Oberbeck for his continued patience, encouragement and most capable guidance while acting as faculty advisor; and for permission to use solutions he has obtained by programming the problem for an electronic computer. I wish to thank Professor C. A. Magwire for his valuable assistance as second reader, and Mrs. D. P. Slingerland for her meticulous preparation of this typescript. Appreciation is also expressed here to personnel of VX-4 and the U. S. Naval Air Missile Test Center, Point Mugu, California, who provided much helpful information on the practical aspects of testing.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

TABLE OF SYMBOLS

(Listed in order of their use in the test)

p_e	an estimate of the parameter, p
p	a parameter value (probability of success)
x	a variable representing the outcome of a trial
n	the number of trials in any testing program
L_1	the lower limit of a confidence interval
L_2	the upper limit of a confidence interval
$\underline{a}$	a confidence coefficient
p_n	specific values of p $(n = 1, 2)$
α	the maximum risk of rejecting an acceptable weapon
β	the maximum risk of accepting an unacceptable weapon
s	the number of successes in a series of trials
$f_1(p)$	uniform probability density function
$f_2(p)$	triangular probability density function
θ	a particular value of p_e which characterizes the terminal decisions
$d_1^{\ t}$	the decision to convert (terminal decision)
$d_2^{\ t}$	the decision to develop (terminal decision)
W	a weight function
∂	a parameter which characterizes W
C	cost function
c	the cost of one trial
N_m	the minimum number of trials required
N_M	the maximum number of trials required
(i, j)	coordinates representing the number of failures and successes observed

CHAPTER I

INTRODUCTION

This thesis is concerned with the application of some statistical methods to testing programs of the type conducted by OpDevFor or similar agencies, operating directly for CNO. Such programs are defined as naval operational testing. The thesis does not consider applications of these methods to quality control or to development or engineering tests such as those which might be conducted by the material Bureaus.

The scope is further limited to those operational testing programs in which the object is to obtain an estimate of the probability of success of a weapon or weapon system, in future combat. As the phrase is used, probability of success can be thought of as generally equivalent to hit probability, percent effectiveness or reliability. The term testing program is used to describe any test which consists of a series of independent trials conducted to obtain the above estimate. The estimate is regarded as providing a quantitative basis for executive decision at some level of the Naval organization.

A fundamental problem in any testing program is the determination of the number of trials to be conducted. For some programs this can be difficult, depending essentially on the magnitude of various "costs". It is clear that the number of trials is related to cost. The reconciliation of all the costs involved is sometimes difficult. This concept is well phrased by Breakwell [3] . In a sentence taken out of context:

> A balance is sought between (1) the cost of testing for reliability and (2) the risks, because of limiting testing, of either accepting an insufficiently reliable product or rejecting a sufficiently reliable one.

As (1) increases the natural tendency is toward fewer trials; as (2) increases the tendency is to desire more trials. Thus, if a balance is to be obtained, the statistical methods used must explicitly relate the "cost", the number of trials and the decisions to accept the product or weapon. Statistical decision theory provides a rational basis for attacking this problem but it can only be applied when these costs can be specified. Frequently, these costs cannot be estimated by the testing agency. Under these circumstances, the costs must be furnished to the agency or the agency is compelled to apply a statistical method which is not based on such estimates. Chapters II and III are devoted to such methods.

Line officers who are unfamiliar with statistical techniques may often be directly concerned with testing programs. Chapter II serves to introduce the technique of confidence intervals, which are often used in reporting test results. It points out that this technique does not provide the planner of the testing program with adequate guidance as to how he should specify the number of trials. Chapter III illustrates the use of sequential analysis. It is considered applicable to testing programs which are conducted to compare improved weapons with an existing weapon.

Testing programs which involve new weapons such as guided missiles, where the costs in (1) and (2) are high, provide a possible field for application of statistical decision theory. It is considered that the

cost estimates required may be available to naval planners at the CNO level. Chapter IV illustrates the planner's role in the application of this theory to the problem of testing a guided missile. This chapter can be read without reference to preceding chapters, if desired.

CHAPTER II

CONFIDENCE LIMITS

Consider the problem of a fleet testing agency in estimating the effectiveness or usefulness of a weapon or weapon system. For example, a destroyer which is equipped with a new or improved anti-submarine weapon. In order to express the effectiveness of this weapon system quantitatively, some measure must be used. Suppose the measure chosen is the percentage of hits achieved by the system in a series of independent trials. This measure is regarded as an approximation of the percentage of hits which will be achieved by the system in a future war but the actual percentage of hits, or the true value of p as it will be called, is an unknown quantity. It is assumed that this true value can be estimated by suitable testing. This estimate will be designated by p_e.

Assume that the naval planner has a testing program which simulates as far as possible the combat conditions under which the system might be used. Also, that the number of simulated attacks or trials will be fairly large (50 or more) and will be conducted so that they represent a random sample of observations. Further assume, that the number of trials to be made is fixed by limitations over which the planner has no control.

It is intuitively apparent that the accuracy of the estimate, p_e, will depend upon the number of trials conducted; the larger the number of trials, n, the greater the accuracy of the estimate will be. If we

designate the result of a trial by x, then we may consider that x can have only two values; $x = 1$ for success or hit, and $x = 0$ for a failure or miss. The estimate p_e may then simply be the total number of hits divided by the total number of trials. However, this estimate may not precisely represent the true p, therefore, a measure of the possible uncertainty in p_e is desirable since the test result will be used as a basis for making statements about the true p of the system. The use of confidence intervals, or limits, provides this measure. This technique is best illustrated by an example.

Suppose 50 trials of the system have been conducted and 15 hits scored; thus $p_e = \frac{15}{50} = .30$. What statements can the planner make about the true value of p? By the statistical method known as determining confidence intervals, two limits, say L_1 and L_2, can be computed, he can then say that the true value of p lies in the interval between these limits, but he can make this statement only with some arbitrary degree of assurance that it is correct. This degree of assurance or confidence is expressed by a confidence coefficient $\underline{a}$. Its value depends upon the degree of confidence the planner desires to have when he makes the statement that p lies in the interval L_1 to L_2. If he wants to be 95% certain then the statement would be

$$\text{(A)} \quad \text{Prob } (L_1 < p < L_2) = .95$$

where $L_1 \doteq .17$ and $L_2 = .43$, for this example. This expression should be read: "The probability is .95 that the variable limits L_1 and L_2 include the true value p between them". This implies that there is a 5% chance of being wrong and that the true value of p

might be outside this interval. A similar statement could be made with, say 99% confidence (one chance in 100 of being wrong), but if this degree of assurance were demanded the effect would be to spread the limits L_1 and L_2 farther apart, that is, (.15 to .49). Thus, the planner would be <u>more assured</u> about the truth of his statement but at the same time <u>less certain</u> of the value of p.

As Mood [4] points out, (A) should be carefully interpreted because it appears that p is a variable when actually it is not, p being a fixed value, the true hit probability of the weapon. The variables are L_1 and L_2. With this fact in mind, (A) has the meaning that we are 95% certain that the interval formed by L_1 and L_2 includes p.

L_1 and L_2 are used to represent the following variables:

$$L_1 = \left(p_e - 1.96 \sqrt{\frac{p_e(1-p_e)}{n}}\right)$$

$$L_2 = \left(p_e + 1.96 \sqrt{\frac{p_e(1-p_e)}{n}}\right)$$

From (A) it can be seen that the limits L_1 and L_2 are functions of p_e, $\underline{a}$ and n, that is, in functional notation

$$L_1 = L_1(p_e, \underline{a}, n)$$

$$L_2 = L_2(p_e, \underline{a}, n)$$

These relations show that the limits depend upon the outcome of the test, p_e, the number of trials, n, and the confidence coefficient, $\underline{a}$. Thus, as pointed out above, if $p_e = .3$, $n = 50$ and $\underline{a} = .95$, then L_1 and L_2 are determined as .17 and .43, respectively.

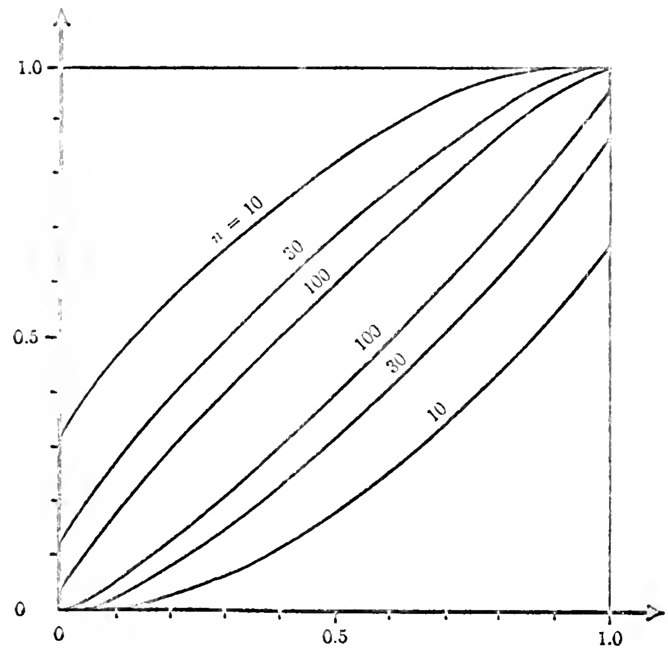
The planner may feel that the confidence interval .17 to .43,

computed on the basis of $p_e = .3$ from 50 trials, is too large and that if more trials had been conducted he could have located p within narrower limits. If he had obtained the same outcome of $p_e = .3$ on tests which had consisted of 50, 100 and 1000 trials, respectively, the following Table illustrates the change in the confidence interval as n and $\underline{a}$ are changed:

No. of trials	Approx. 95% limits	Approx. 99% limits
50	.17 - .43	.15 - .49
100	.21 - .40	.19 - .43
1000	.27 - .33	.265 - .335

Figure 1 is a chart which illustrates the dependence of the confidence interval on n and p_e. For a given value of p_e, a vertical line intersects two curves corresponding to a given value of n. These intersections, projected on the vertical axis, are the limits L_1 and L_2; and the interval formed by these limits spans the true value of p. Hence, it is labelled as the p axis. This chart is for the confidence coefficient of .95 and clearly shows the effect of increasing n. Also the number of trials, n, may be regarded as a function of the interval $[L_2 - L_1]$, $\underline{a}$ and p_e. That is, all three quantities must be known in order to determine n. In functional notation.

$$n = n(L_2 - L_1, \underline{a}, p_e)$$



p_e

Confidence Interval Chart for
Confidence Coefficient $\underline{a}$ = .95

Figure 1

It should be noted that the curves are not very useful for attempting to determine, in advance, the number of trials required to give a fixed confidence interval of desired length because this would mean the outcome of the testing program, p_e , would have to be known in advance.

It is apparent that the use of confidence limits has a place in naval testing as a means of stating the results of a testing program in a precise manner which is more meaningful than simply stating the outcome of the program as a single number p_e . But the use of this technique does not provide adequate guidance for advance planning to indicate how

many trials should be run or what degree of confidence should be stipulated. This technique is based on a fixed number of trials. The planner may have chosen the number, in advance, from considerations of time, services required, etc., but we wish to emphasize that this theory does not provide any criteria upon which the planner can base such a choice.

CHAPTER III

SEQUENTIAL ANALYSIS

Sequential Analysis was developed by A. Wald in 1943 for use on problems which arose during World War II. It was widely used in manufacturing establishments for acceptance inspection of "lots" of mass production items. The detailed application of sequential tests to such problems is given by the Statistical Research Group [5]. The principal advantage of sequential tests in acceptance inspection is that it reduces the amount of inspection required. As shown in [5], the methods of sequential analysis can be applied to experiments. Since certain types of fleet testing problems can be thought of as "experiments", the application of sequential tests to a testing program will be described.

Sequential Analysis can be used when a testing program is to be conducted for the purpose of comparing the hit probability or the probability of success of a supposedly improved weapon system with that of the existing system. Testing of this type may be indicated when it is desired to use the test results as a basis for decision to recommend acceptance or rejection of the modified system, for fleet use.

Sequential analysis does not permit the exact number of trials required to be determined in advance, however, an average or expected number may be calculated. From a naval planning standpoint, ignorance of the total number of trials required may pose some problems for scheduling, determination of material requirements, services and related details. This may be a disadvantage, but the use of sequential tests can

result in a possible economy of trials required to reach a decision. This economy may represent considerable savings of time and services.

In order to use Sequential Analysis, the planner must be able to specify certain quantities or inputs. The following hypothetical example will illustrate these inputs and how they might be specified. The example will be similar to testing problems of the type faced by fleet testing agencies. The test results, which might have been obtained for this hypothetical example are shown in Table I.

EXAMPLE: A destroyer equipped with a supposedly improved anti-submarine weapon is to be tested to determine the hit probability of this system (designated as system II). The hit probability of system II is to be compared with the known hit probability of otherwise identical destroyers employing a weapon which has been in service use (designated as system I). We shall assume the hit probability of system I to be .2 . Also, we shall assume that the cost of the two systems is approximately the same.

Assume finally, that it is desired to specify a sequential test. The outcome of testing will indicate whether to accept or reject system II as being better on the basis of the sample of trial runs. Therefore, the inputs of the Sequential Test must be carefully specified. These inputs uniquely define the sequential test:

(a) p_1 - The hit probability which would make the new system II "Unacceptable".

(b) p_2 - The hit probability which would make the new system II "Acceptable".

(c) α - The maximum allowable risk or probability of rejecting a new system II if it has hit probability p_2 or better.

(d) β - The maximum allowable risk or probability of accepting system II if it has a hit probability of p_1 or less.

SPECIFICATION OF (p_1, p_2, α, β) In this example, the first input, p_1, is probably the easiest to select. Since p_1 is the hit probability of a new system which would make it unacceptable, it seems logical that any new system would not be desirable if it were no better than the existing one, that is, if its hit probability was no better than that of system I, which we have assumed to be about .2. Hence set $p_1 = .2$.

The second input, p_2 might be selected by reasoning as follows: At first thought it would appear that any system with a hit probability greater than that of existing systems is "acceptable"; however, since the test will consume time and money it would not be logical to accept a system with a hit probability only slightly greater than existing systems; say, p between .2 to .3 . On the other hand it might be argued that the new system should be at least twice as good as the old to justify expense of conversion, that is, an increase of p to .4 would justify the expense of tests and installation of the new system if it were accepted.

The inputs α and β represent the probabilities of making a wrong decision and these risks are unavoidable. The values of α and β are small and not necessarily equal.

α is the probability of rejecting the new system if is has a hit

probability of $.4$. Since it would be undesirable to reject a new system which is, on the average, twice as good as the existing one, then the probability of making such an error should be made very small. If a risk of one chance in 100 can be tolerated then α would be $.01$.

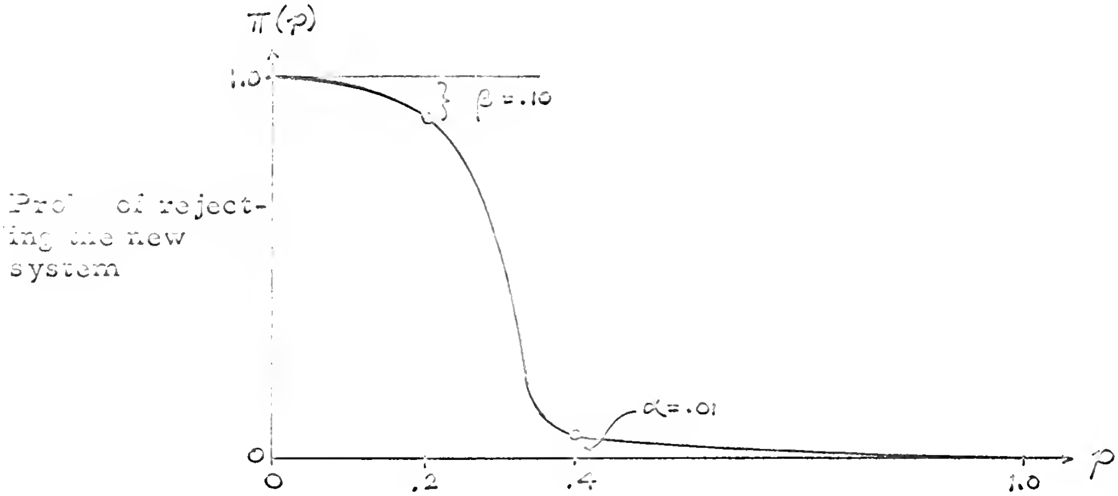
β, on the other hand, is the probability of accepting a new system if it has a hit probability p_1. Since it is possible for the test to lead us to such a wrong decision it would mean that we were accepting a new system which was actually no better, perhaps worse, in terms of hit probability than the existing one. Hence, we want to make the probability of making such an error small also. But, we can tolerate a greater risk of this error than we can of rejecting a better system, so β can be made larger than α. In other words, accepting system II when it has the same hit probability as the old system is not too serious from a military standpoint. Therefore if β were selected as $.10$, this would be taking one chance in ten of making such a wrong decision. For illustrative purposes, choose $\alpha = .01$ and $\beta = .10$.

Since either error is possible, it would seem desirable to have the risks, α and β, of making such errors as small as possible. That is, make α and β even smaller than the values chosen above. It will be seen that demanding smaller risks will result in having to make a greater number of trials and this number can become unacceptably large; therefore some risk must be tolerated to avoid a prohibitive number of trials.

Having selected the inputs

$$p_1 = .2 \qquad p_2 = .4 \qquad \alpha = .01 \qquad \beta = .10$$

it is mathematically possible to determine what is defined as a Power Curve, $\pi(p)$. This curve will represent the probability of rejecting system II as a function of p. It will have the following general shape:



Power Curve for Case A

Figure 2

The ordinate at any point $p = p'$, represents the probability, $\pi(p')$, of rejecting the new system when its hit probability is p'. At $p = p_2 = .4$, $\pi(p)$ has the value of $\alpha = .01$, which is the probability of rejecting system II when its hit probability is $.4$. At this point we are taking one chance in 100 of rejecting system II when its true p is equal to $.4$. Note that if system II has a $p > .4$, we have a still smaller probability of rejecting it.

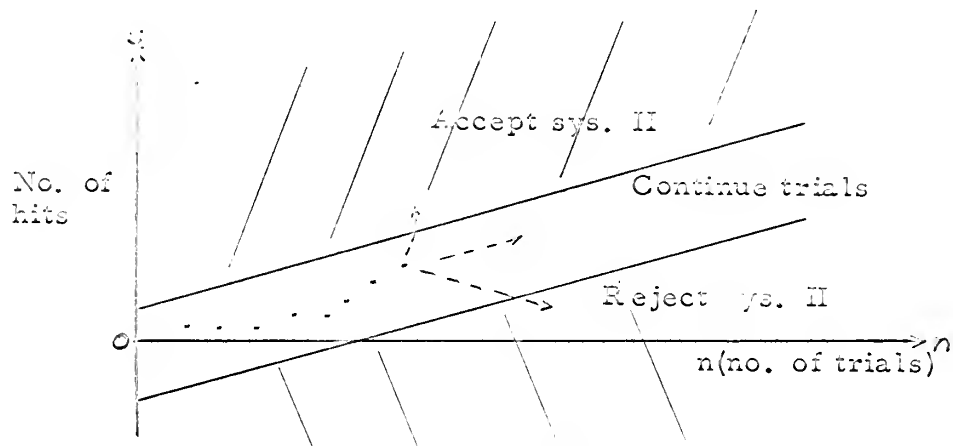
At the point $p = p_1 = .2$ we have a very high probability of rejecting the new system and since (one minus the probability of rejection) is equal to the probability of acceptance, then $[1 - \pi(p_1)] = \beta = .10$.

This is the risk we are willing to take in accepting the new system when it is no better than the existing one.

Between p_1 and p_2 the new system will be rejected with probilities varying from $1 - \beta$ to α - the probability of rejection decreasing as we approach the "acceptable" hit probability of .4 .

The Power Curve need not actually be produced in order to make use of sequential testing.

USE OF THE TEST The four inputs, p_1, p_2, α, β are used to construct a graph which is the basis of the sequential test:



Graph for Sequential Test .

Figure 3

The graph is easy to use. Assume that the results given in Table 1 are being obtained from a testing program where the outcome of each trial is denoted as success or failure. After each trial, the total number of successes is plotted against the total number of trials conducted thus far. As trials progress the plotted point will either fall in one of the shaded regions labelled Accept System II and Reject System II; or it will

be in between the parallel lines. Trials are continued as long as the point remains between the parallel lines but eventually one of the shaded regions will be reached. It is this uncertainty as to when one of these regions will be reached that precludes advance determination of the number of trials required.

The details of constructing the straight lines which comprise the boundaries of the three regions is given in Appendix A .

Figures 6, 7 and 8 are constructed using the same test results given in Table 1, for three different sets of input parameters, and will be described as case (A), (B) AND (C).

Trial number	1	2	3	4	5	6	7	8	9	10	11
Outcome	F	F	S	F	S	S	F	F	S	S	F
	12	13	14	15	16	17	18	19	20	21	22
	F	F	F	S	F	F	S	S	F	F	F
	23	24	25	26	27	28	29	30	31	32	33
	F	F	F	F	F	F	S	S	S	S	S

Table of Assumed Test Results
Table 1

CASE A

Figure 6 illustrates the results that would have been obtained using the values

$p_1 = .2 \qquad p_2 = .4 \qquad \alpha = .01 \qquad \beta = .10$

Note that acceptance of System II as having a hit probability .4 or greater, would have occurred at the 19th trial where the plotted point crossed the boundary into the acceptance region.

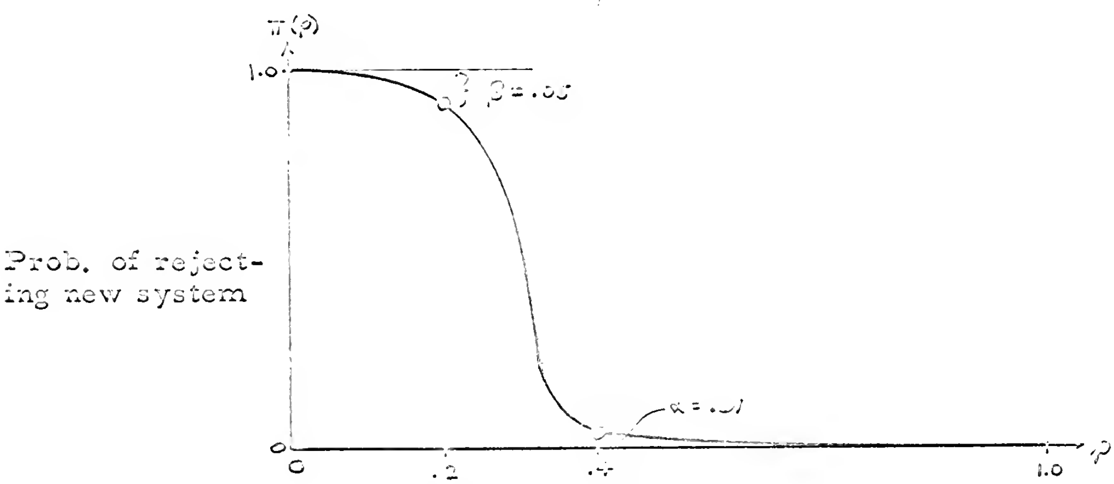
The Power Curve for this particular set of input parameters has been shown earlier.

CASE B

Input parameters:

$p_1 = .2 \qquad p_2 = .4 \qquad \alpha = .01 \qquad \beta = .05$

Here, the power curve will appear much the same as for CASE (A) but note that β is now smaller.



Power Curve for Case B

Figure 4

Decreasing β from .10 in Case (A) to the value of .05 in this case, is equivalent to saying that the planner wants a smaller risk of accepting System II if it is no better than System I. A decrease in the risk might be considered necessary because of the economic aspects,

such as the desire not to take too great a risk of changing to a new system which is in reality no better than the existing one, if the change represents considerable expense. This is equivalent to desiring more protection against the risk of a wrong decision and will result in more trials being required.

Figure 7 shows the test data plotted. The boundaries are different from Case (A), as indicated by the equations of the lines.

Note, the acceptance of System II results at the 33rd trial as opposed to the 19th trial in Case (A). This increased number of trials is the result that was anticipated. The number of trials, for the smaller risk β, is greater.

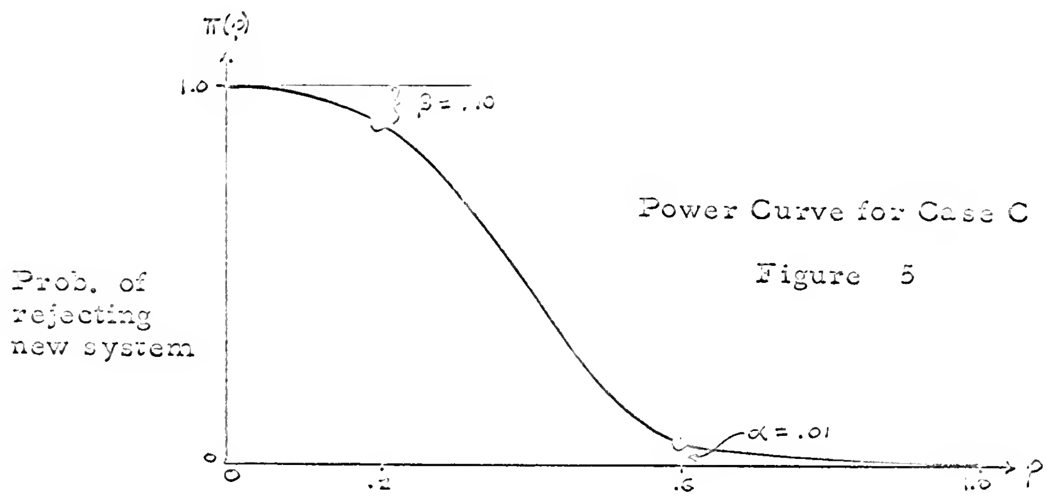
CASE C

Input parameters:

$$p_1 = .2 \qquad p_2 = .6 \qquad \alpha = .01 \qquad \beta = .10$$

Before discussing test results in this case, it is important to note the implications of setting $p_2 = .6$.

The power curve will have the following general shape



Power Curve for Case C

Figure 5

Here, p_1 and α are the same as in Case (A) but the planner in now saying he will accept the new weapon on the assumption $p_2 = .6$ and still take the same risks $\alpha = .01$ of rejecting the new system for $p = .6$. Note from the power curve that if the true p of System II is in the neighborhood of .4 or .5 he is taking a greater chance of rejecting System II than in Case (A). This is just another way of saying that he is not too concerned with the interval between .2 and .6 so the test will reject systems which have a true hit probability in this region with a higher probability than in Case (A). Or, System II must have a higher hit probability, on the average, than in Case (A), to be accepted by this test.

Figure 8 shows the result of the test data. The boundaries are again different than those of Case (A) and (B).

Note, that on the 28th trial the sequential test indicated rejection of System II as not having a hit probability of .6 or greater.

It is important to recall that in any of these three cases there is always the chance that a wrong decision will be made. We have tried to keep the chance or risk of such eventualities small, namely by keeping α and β small, consistent with a reasonable number of trials. A demand for less risk of wrong decisions will result in more trials being required. Further, α and β had to be selected on what might be called an "intuitively reasonable" basis. The sequential test does not provide any means whereby the cost of testing and cost of wrong decision can be reconciled.

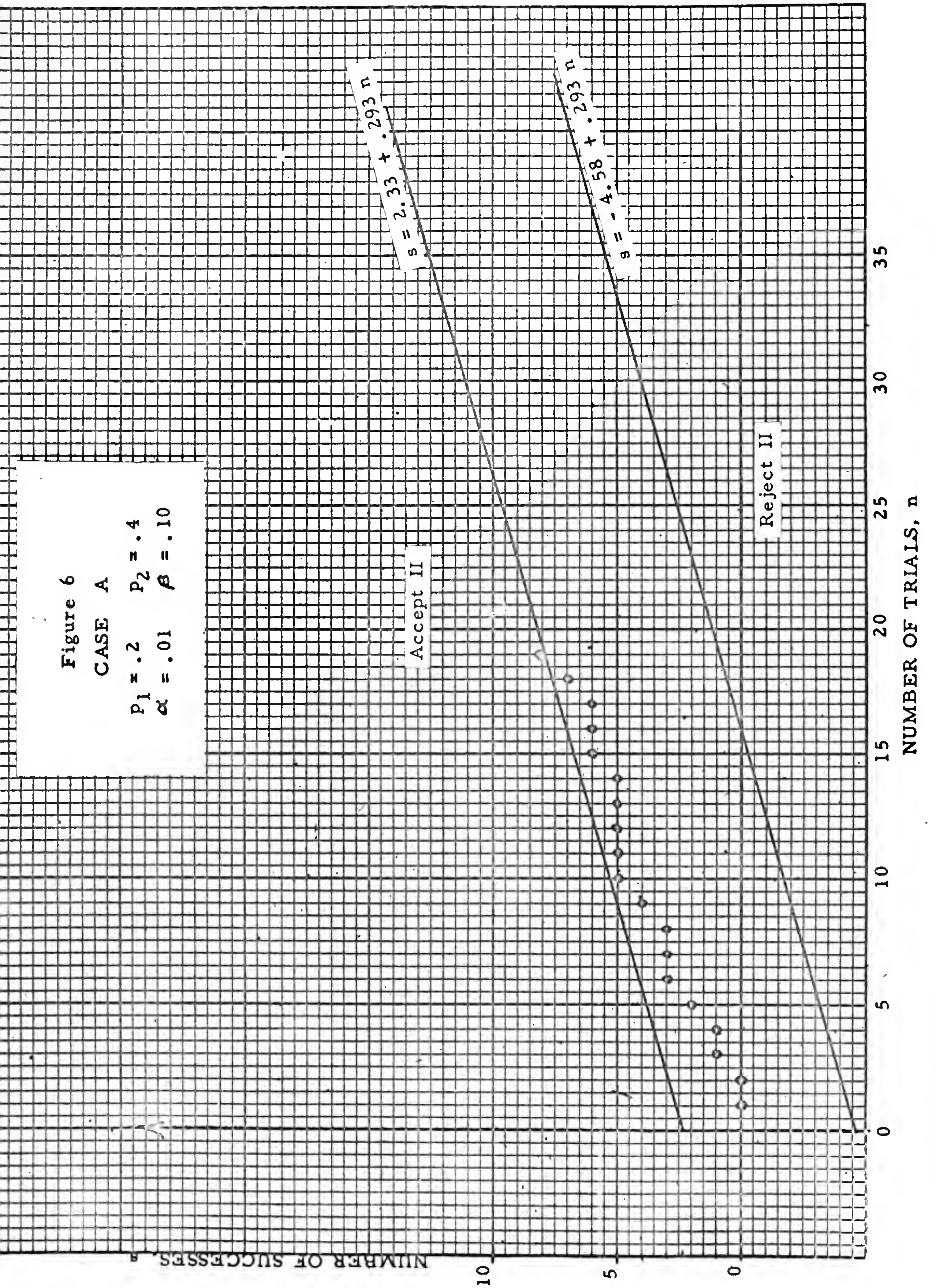


Figure 6
CASE A
$p_1 = .2 \qquad p_2 = .4$
$\alpha = .01 \qquad \beta = .10$

Figure 7
CASE B
p_1 : .2 p_2 : .4
α : .01 β : .05
s = 3.03 + .293 n
s = - 4.64 + .293 n
Accept H
Reject H
0
5
10
15
20
25
30
35
NUMBER OF TRIALS, n

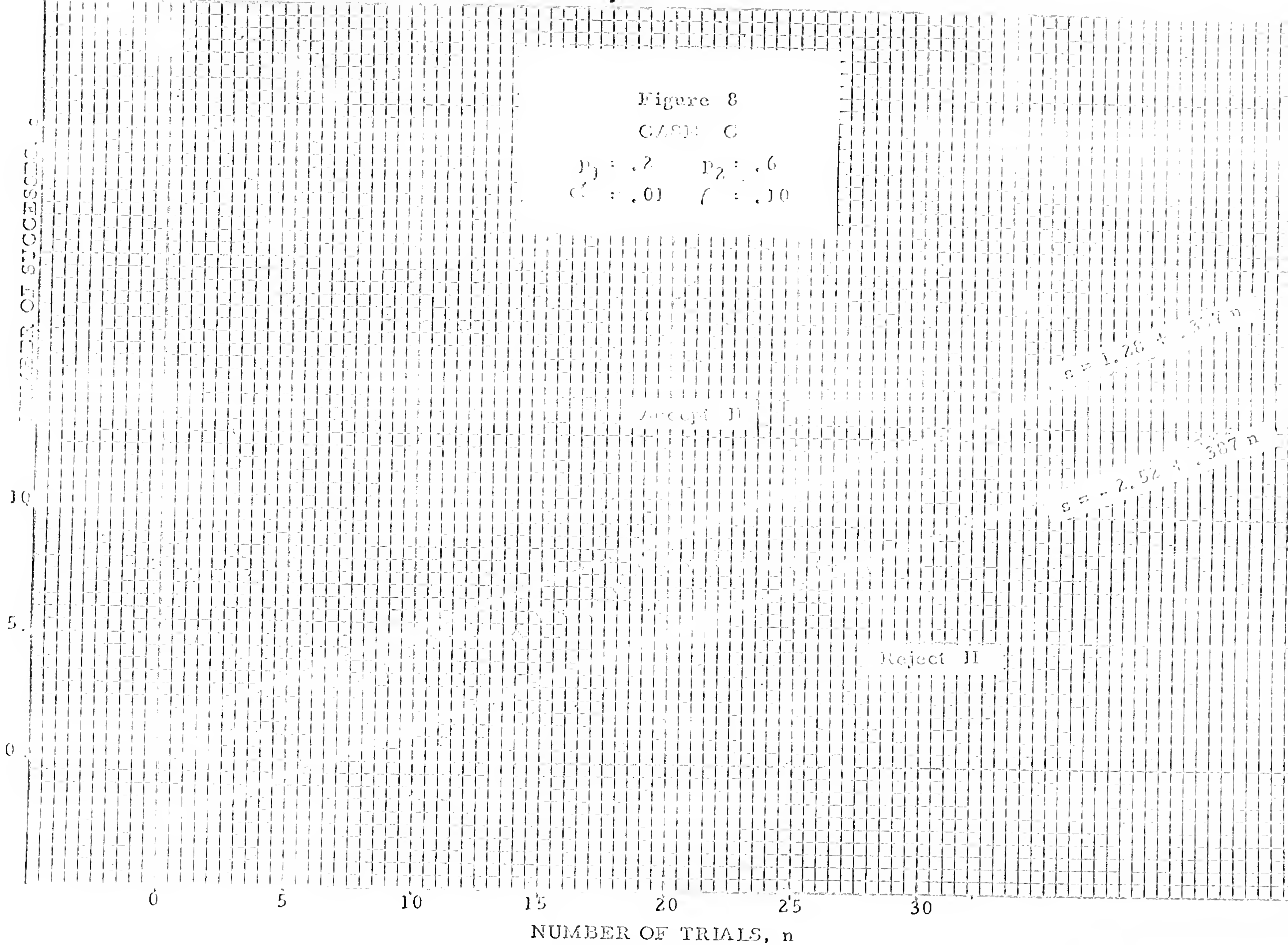


Figure 8
CASE C
p_1 : .2 p_2 : .6
α : .01 β : .10
Accept H
Reject H
s = 1.28 + .387 n
s = - 2.52 + .387 n
0
5
10
0
5
10
15
20
25
30
NUMBER OF TRIALS, n

Implications of Sequential Testing.

1. In the three cases A, B and C, it should be noted that the terms acceptance or rejection are used with reference to assumptions or hypotheses about the new system. In other words, a statistical test is the test of an assumption. In cases A and B we tested the hypothesis that System II had a hit probability of .4 and as a result of the tests we accepted this hypothesis. In Case C, rejection of the System II was indicated.

There could be reasons for deciding not to use a new system even if the statistical results indicated that it was desirable. For example, the new system may be too heavy, or useless in rough seas, or require excessive maintenance. If many such factors weighed against a new system, obviously it would not be acceptable.

2. The example given is just one illustration of the use of sequential tests. They are not limited to trials to determine hit probability (Binomial distributions) but could also be used in trials where, for example, the miss distance between weapon and target is being measured. While the details of the example used in this chapter would not apply to such a test, the principles are the same. See [5].

3. It has been mentioned that advance planning may be handicapped because the exact number of trials cannot be specified but this fact should not preclude the use of Sequential Analysis. In many situations, the knowledge of the expected number of trials required may be sufficient for planning. The great improvement of Sequential Analysis over the classical test procedures, namely, the reduction in the number of

trials required to reach a decision, should not be overlooked.

4. The risks α and β are inputs to the problem which are based upon economic and military considerations. The concept of cost was implied when a value of hit probability of an "acceptable" system, p_2, was chosen. Also, when the corresponding risk α of rejecting such a system is specified, there is an associated idea of cost. Further, there is the cost of testing program which might include services, material, cost of the weapons expended in test, etc. These costs are mentioned to point out that, although the planner considers them when specifying the inputs, the theory does not permit the planner to explicitly take such costs into account.

A more general theory is described in the next chapter. Explicit expressions for cost of testing and costs of wrong decisions provide a means for determining, prior to testing, the number of trials to be conducted.

CHAPTER IV

STATISTICAL DECISION THEORY

This general theory may be useful to naval planners at the CNO level. In order to discuss the application of the theory it seems appropriate to describe a situation giving rise to a Statistical Decision Problem.

(A) EXAMPLE

1. A need for an Air-to-Air Guided Missile is recognized and translated to an Operational Requirement by an office of CNO

2. The Operational Requirement passes to the cognizant Bureau. A contract is let and a missile is produced. At this point testing by Bureau and contractor is involved, but this paper is not concerned with these tests. At some future time the Bureau will inform CNO that a missile has been produced to satisfy the Operational Requirement.

3. It is assumed that CNO now desires an estimate, p_e, of the percentage effectiveness, or probability of success, of the new missile in future combat. In order to obtain this estimate, missile testing by some fleet agency, such as OpDevFor, is indicated. An office of CNO must decide how many weapons should be tested (that is, how many trials should be conducted) so that the testing agency may plan the testing program. This question of the number of trials required constitutes the Statistical Decision Problem.

Statistical Decision Theory provides a rational basis for answering such questions, moreover, the answer is provided prior to testing.

which is advantageous from the point of view of planning. In this chapter we shall examine the form of the answer provided by this theory in a special set of circumstances which will be related to the problem posed in the above example. This special set of circumstances includes:

(a) In the trials from which the estimate p_e is obtained, the outcome of any trial must not be affected or influenced by the outcome of preceding trials, i.e., the trials must be statistically independent.

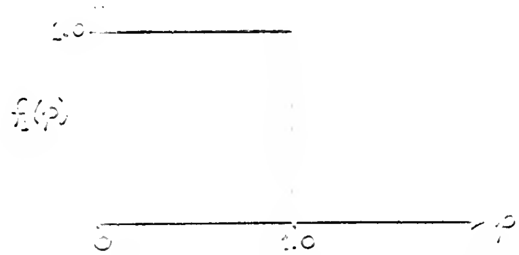
(b) Probability of success of the missile is defined as the probability of successful launching, guiding and detonation of the missile at the intended point. In actual combat there will exist some value for this probability of success, which we will call the true value of p. This value is an unknown quantity but can be estimated by testing.

(c) The outcome of each trial is valued as one for success and zero for failure. Note that failure here is not concerned with which of the three phases of (b) fails. In any series of independent trials, an estimate p_e of the true probability of success p is given by the number of successes divided by the total number of trials.

When these special circumstances are fulfilled, then an answer is given in the form of tables presented at the end of the chapter. The use of these tables requires a specification of certain inputs to the problem. These inputs will be explained in the light of the hypothetical testing problem.

Input I

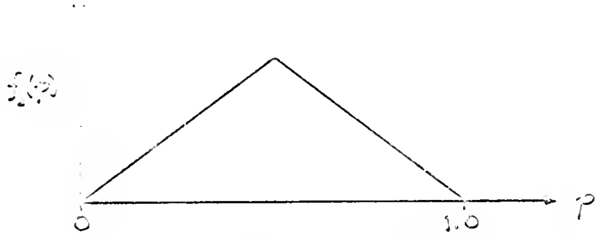
This input represents the concept that the true value of p , since it is a probability, can only take on values between zero and one. Hence it can be visualized as any point in the interval [0, 1] . In this example it is assumed that all values of p in this range are equally likely. This merely says that the planner has no a priori knowledge favoring some values of the true p above others. In technical language, he has specified a uniform probability density function, $f_1(p)$, for p . Graphically:



Uniform Probability Density Function

Figure 9

On the other hand, if the planner had some a priori knowledge which would lead him to favor certain values of p , he might specify a probability density function which reflects this knowledge, such as:



Triangular Probability Density Function

Figure 10

This indicates a knowledge that values of p in the neighborhood of one-half are more likely than others.

When the planner specifies an a priori density function of p as we have done, the solution we will obtain to the Statistical Decision Problem is known as a Bayes solution. If an a priori density function is not specified, the solution, if any, is termed a Minimax solution. Minimax solutions will not be considered in this thesis. An elementary discussion of these solutions is given by Tucker [1], and a more advanced treatment by Wald [2].[1]

<u>Input II - The Final or Terminal Decisions</u>

The estimate, p_e, will be the basis for making decisions concerning the missile. Hence, it is necessary to specify what action is taken after a value of p_e has been determined. Action will result from what we shall call terminal decisions.

In order to apply the theory, the planner must be able to specify a number which we shall refer to as θ in the interval $[0, 1]$. This value of θ is used to specify two terminal decisions.

(1) $p_e > \theta$ will lead to a decision to adopt the missile and will imply that fleet units will be converted to the use of this missile. We shall designate this first terminal decision as the <u>Decision to Convert</u> or briefly d_1^t.

(2) $p_e < \theta$ will lead to the decision to order further development of the missile. This decision would imply that the missile has shown promise of being feasible and there is a reasonable expectation of improving missile performance to a point

[1] Chapter [illegible]

that would warrant its adoption for fleet use. We shall designate this second terminal decision as the <u>Decision to Develop</u>, or d_2^u .

The value assigned to θ might be obtained by a comparison of the effectiveness of existing interceptor A/C weapons and determining what value of probability of success, <u>if</u> attainable by the new missile, would represent at least satisfactory or acceptable interceptor effectiveness.

Suppose θ were chosen as .5 . This would be equivalent to stating that a missile with a probability of success of .5 will result in acceptable or satisfactory interceptor effectiveness against enemy bomber capabilities for the immediate future. Further, it should be a value which is <u>reasonable</u> to expect in view of missile performance during the course of development thus far.

<u>Definition of Wrong Decisions</u>

Having defined the terminal decisions it is now important to note that the test result, p_e , can lead to a wrong decision in two ways. To understand this, the difference between p_e and p must be kept in mind. We would like to think that p_e was always an "accurate" estimate of p , but such is not the case. The p_e indicated by the tests can be greater or less than the true value of p by an unknown amount. The two cases which lead to wrong decisions are shown graphically:

(See following page for figure)

(a) (b)

Cases Leading to Wrong Decisions

Figure 11

Figure 11 (a) is the case where the test result, p_e, is greater than θ while, in fact, the true value of p is less than θ. In such a case the Decision to Convert would be made. It would be a wrong decision. In (b), the test result, p_e, is less than θ while, in fact, the true value of p is greater than θ. In such a case the Decision to Develop would be made and it would be a wrong decision.

<u>Input III The Cost of Wrong Decisions</u> (Weight Function)

It is through the Weight Function that cost of wrong decision is introduced into the problem. Before specifying the form of this function it is necessary to understand the nature of these "costs" in this example. First, they do not represent cost in the sense that money must be paid out. They represent the "value" of the possible consequences suffered as a result of a wrong decision. Reflection will show that this definition can lead to situations which cannot be measured in terms of dollars. For instance, suppose as a result of a test the missile is "accepted", i.e., Decision to Convert is made. Suppose a war ensues and we find our interceptor weapon performance almost nil (That is, the true p is very low). This could lead to loss of a Task Force due to lack of protection - the Task Force loss could mean delay or

inability to execute the overall mission. Obviously, it is practically impossible to assign a dollar value to such consequences. If this could be done it would be taken into account in the theory.

Rather than attempt to carry the evaluation of losses due to wrong decisions to such length, it is considered reasonable and practical to think of the "cost" of a wrong decision as an <u>erroneous expenditure of funds</u>. By so doing the planner will be evaluating the losses or cost of wrong decisions in terms of the dollar budget under which he must operate. These losses can now be associated with consequent cost of either of the two terminal decisions.

1. When the Decision to Convert is made, the cost of this decision may include:

   (a) Cost of converting fleet aircraft, and fleet units to use the missile, that is, cost of structural changes, necessary fire control systems, etc.

   (b) Existing aircraft may not be suitable and new aircraft, designed around this missile and its fire control system, may be needed.

   (c) Cost of actual missiles required by the fleet for some subsequent budgetary period, that is, mass production costs.

   Let us assume these costs to be on the order of $10,000,000.

2. When the Decision to Develop is made, the cost of this decision will include:

   (a) Cost of additional development work by contractor and services incident to the work to develop the missile further. This will represent some fraction of the total devel-

opment funds available.

Let us assume that this cost is also on the order of $10,000,000. The theory does not require the costs of (1) and (2) above to be equal but the solution is simplified if they are equal.

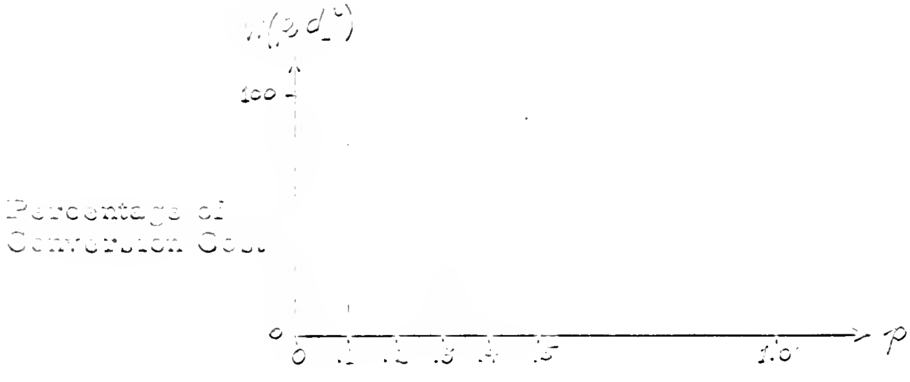
The writer has no information on how funds to cover such cases as listed above are identifiable in the military budget. It is assumed that there are categories of funds which are either ear-marked for fleet improvements of the nature of those in (1), or development work as contemplated by (2). The main point is that definite costs must be associated with these decisions, and the planner must be able to provide reasonable estimates of these costs.

Linear form of the Weight Function

The form of the Weight Function describes the manner in which the planner will penalize or assess himself when a wrong decision is made.

Refer to Figure 11 (a). When the Decision to Convert is a wrong decision, we would be spending conversion funds for a missile which did not meet our criterion ($0 = .5$) for acceptable performance. One way to penalize ourselves for this wrong decision would be to make the penalty proportional to the amount that the missile fails to meet the criterion. For instance:

True value of p	Penalty
.5	0
.4	20% of conversion cost
.3	40% " "
.2	60% " "
.1	80% " "
0	100% " "



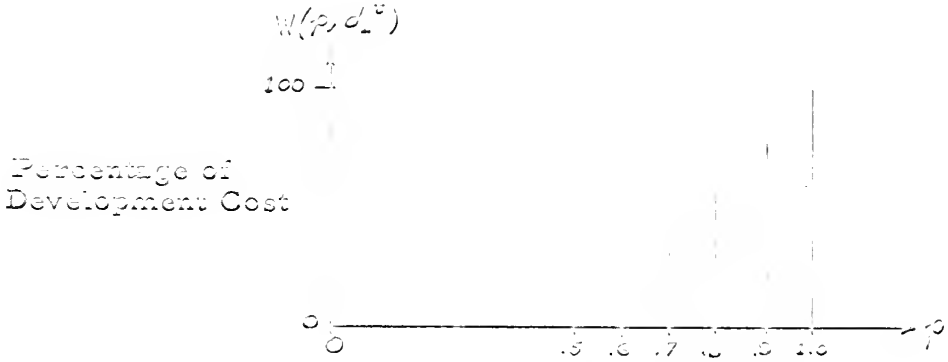
Penalty Associated with d_1^u

Figure 12

This graph shows what we can call the cost of wrong decision associated with d_1^u . It is defined as a Weight Function, W. Symbolically: $W = W(p, d_1^u)$. It indicates that we penalize outselves nothing if the true $p > .5$ since a missile with a value of p in this range is as good or better than we asked for. At $p = .4$ we have a missile almost as good so the penalty is not very great. Similarly, the smaller the value of p , the greater the penalty for spending conversion funds.

By similar reasoning we can define a cost for the other wrong decision which can be made. Refer to Figure 11 (b) . When the Decision to Develop is a wrong decision, money is spent to develop a missile which is already as good or better than the criterion $(\theta = .5)$. We could therefore penalize outselves in proportion to the amount that p <u>exceeds</u> the criterion.

True value of p	Penalty
.5	0
.4	20% of development cost
.3	40% " " "
.2	60% " " "
.1	80% " " "
0	100% " " "



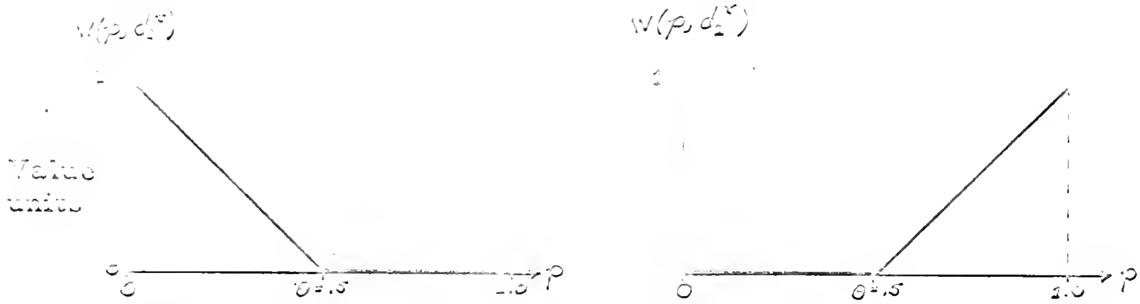
Penalty Associated with d_2^u

Figure 13

This is the graph of the cost of the second wrong decision that can be made or $W = W(p, d_2^u)$. There is no penalty for $p < .5$ because a missile with a value of p in that range requires further development, according to our criterion. But for a value of p , say $p = .75$, some penalty should be assigned since a missile with this value of p does not require further development. Similarily, as p increases the penalty increases. By this process, the greater the value of p , the greater the penalty for spending development funds.

If the increments in Figure 12 and 13 are imagined to get smaller and smaller the discrete function will approach a straight line. This is equivalent to defining a linear Weight Function. To simplify the mathe-

matical details the costs of conversion and development ($\$10^7$) are assigned the weight of one value unit.



(a) Case where Decision to Convert is made

(b) Case where Decision to Develop is made

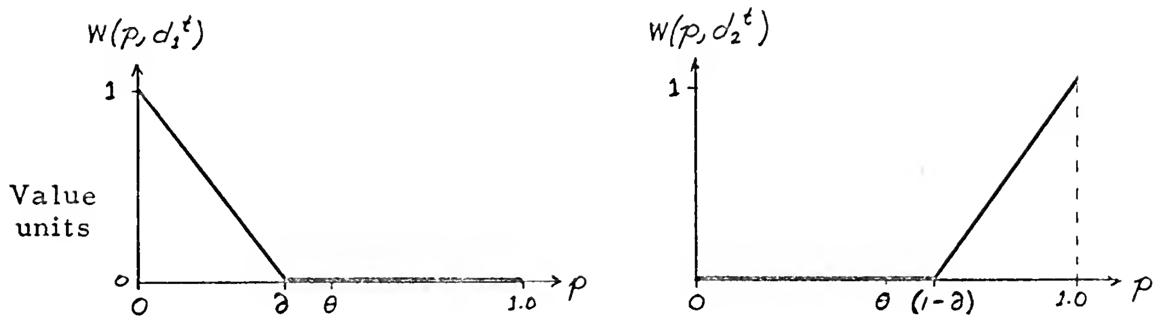
Linear Weight Function

Figure 14

Consider what Figure 14 (a) indicates. There is some penalty for every value of $p < \theta$. Suppose p is in the interval .4 to .5. A missile with p in this range is almost as effective as one with a p of .5. Consequently, when the decision to convert is made, it seems logical that there should be no penalty in cases where values of p are only slightly less than θ. Precisely, we can define an interval $[\delta, \theta]$ where $W(p, d_1^*) = 0$.

For Figure 14 (b), similar reasoning will apply. Suppose p is in the interval .5 to .6. A missile with p in this range requires only slightly less further development than a missile with p of .5. So, for values of p slightly greater than θ, it is also logical that there should be no penalty for the decision. Or, define an interval $[\theta, 1 - \delta]$ where $W(p, d_2^*) = 0$.

The Weight Function can now be graphed as:



(a) Case where Decision to Convert is made

(b) Case where Decision to Develop is made

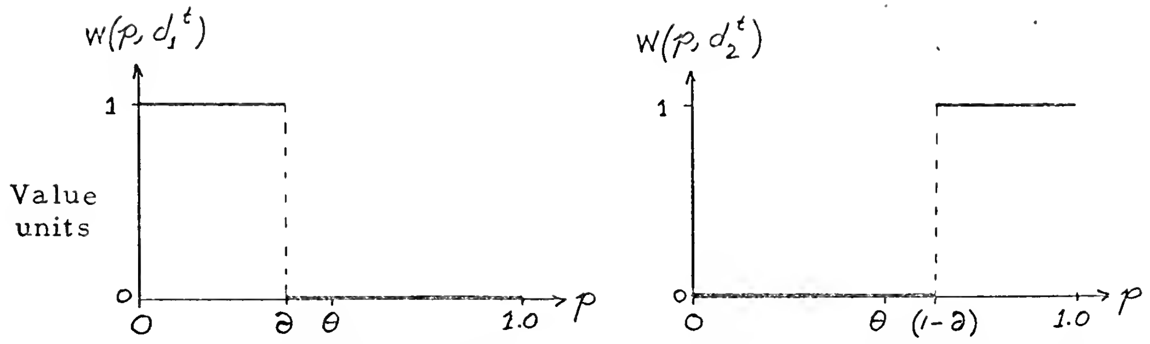
Linear Weight Function
Showing the Parameter ∂
Figure 15

The value of ∂ is at the discretion of the planner. For this example, ∂ is chosen as .33 .

Another possible form of Weight Function

It should be noted that the linear form of the Weight Function resulted from the way in which it was assumed that the cost of wrong decisions could be assigned. As Wald [2] points out, the Weight Function is the most difficult input of the Statistical Decision Problem to determine. He further points out that a Simple Weight Function is suitable for many practical problems.

A simple Weight Function has only two values, zero or one.



(a) Case where Decision to Convert is made

(b) Case where Decision to Develop is made

Simple Weight Function

Figure 16

The meaning of θ , p , ∂ , d_1^t and d_2^t are exactly the same as before.

The difference between this Simple form and the Linear form is that the cost of wrong decision is not proportional to the difference between p and θ . There is no penalty to be assessed for a missile whose true probability of success lies between ∂ and $(1 - \partial)$. But, when the decision to convert is made and $p < \partial$, the penalty is the entire cost of conversion. Similarly, when the decision to develop is made and $p > (1 - \partial)$ the penalty is the entire cost of further development.

The Simple Weight Function has been illustrated so that comparisons of solutions using Simple and Linear Weight Functions can be made.

Input V Cost of testing of Cost Function. (C)

This input is not as difficult to specify. In many testing programs

the cost of each trial in the program is the same. When this is the case, then the cost of testing is proportional to the number of trials.

As in the other costs, the naval planner must be able to estimate this cost of testing. This would include such items as the cost of the missile itself; contractors services required for the tests, etc. C is therefore determined by a cost analysis of charges directly applicable to the testing program for a specific missile. It is not considered that the cost of operating the testing agency for the duration of the testing program is a cost which would be included here.

Suppose the cost of each trial for the missile in this example is on the order of \$50,000. As it is used in Statistical Decision Theory, C is expressed in the same value units as the Weight Function. For the Weight Function with the maximum value of $\$10^7 = 1$ value unit, then define $c = C/W = .005$ value units.

The effect of varying the ratio C/W will be discussed in the next chapter.

(B) Solution of The Example

The preceding discussion has indicated how the inputs required for a Bayes Solution of a Statistical Decision Problem may be specified.

The solutions are given in the form of Table 2, using the Linear Weight Function and Table 3, using the Simple Weight Function.

By way of summary of the problem to which the Tables are solutions, we shall briefly review the special circumstances required of the testing program and the four inputs to the problem.

(a) Recall from Page 26 that a suitable testing program will con-

sist of independent trials, the outcome of whi is either suc- cess or failure. The estimate p_e is the nu er of successes divided by the total number of trials.

(b) Input I. The specification that the true p he missile can take on any value in the interval [0, 1] and at all values of p are equally likely.

(c) Input II. The final or terminal decisions an eir relation to the preassigned value of $\theta = .5$.

d_1^t - the Decision to Convert

d_2^t - the Decision to Develop

(d) Input III. The cost of wrong decision or We Function (W); both Linear and Simple forms.

$W(p, d_1^t)$ - Cost or penalty when d_1^t is a ng decision

$W(p, d_2^t)$ - Cost or penalty when d_2^t is a ng decision

The maximum value of W is assumed to be ^{7} (1 value unit) and $\theta = .33$

(e) Input IV. The cost of testing or Cost Functi (C)

Testing was assumed to cost \$50,000 per tria nce,

$c = \$50,000/\$10^7 = .005$.

Note that c has the same value as long as t atio C/W is the same, i.e., if we had assumed W to be \$ and Cost per trial of \$5,000 then c is again .005.

<u>Explanation of Table 2 (Linear W)</u>; page 40

Any cell in the table can be identified by a pair of bers or coor- dinates, say (i, j), where i denotes the number of ures and j

NUMBER OF SUCCESSES

NO. OF FAILURES	0	1	2	3	4
0	.16667 .02466	.03704 .01966	.00926 .00926	.00247 .00247	.00069 .00069
1	.03704 .01966	.09259 .02546	.02963 .02046	.00960 .00960	.00314 .00314
2	.00926 .00926	.02963 .02046	.05967 .02426	.02254 .01926	.00840 .00840
3	.00247 .00247	.00960 .00960	.02254 .01926	.04138 .02206	.01724 .01706
4	.00069 .00069	.00314 .00314	.00840 .00840	.01724 .01706	.03000 .01835

Solution Using Linear Weight Function

Table 2

the number of successes which have been observed in $(i + j)$ trials. The numbers i and j are the numbers designating each row and column respectively. Thus, row two and column three locate a specific cell after trials which yield two failures and three successes.

Every cell contains an upper and lower entry. Each entry represents a cost expressed in value units. If desired, these entries may be converted to dollars by multiplying by the dollar equivalent of a value unit ($\$10^7$ in this example) . The upper entry is the expected cost (in value units) if no further trials are made and a Terminal Decision to Convert or Develop is made on the basis of trials thus far. The lower entry is the expected cost (in value units) if trials are continued and a Terminal Decision to Convert or Develop is based upon the result of further trials.

The relation between the upper and lower entries means that, in any cell where the lower entry is smaller than the upper, the expected cost is less to continue testing than it is to make a decision at this point. Notice that the cells in which the lower entry is smaller than the upper are enclosed by the dotted line.

It is this dotted boundary which enables the planner to determine, in advance of testing, the limits on the number of trials required. The Table actually yields the minimum as well as the maximum number of trials that might be required. These maximum limits are of interest to the planner. For our example the minimum number of trials, N_m , is 2 and the maximum number, N_M , is 9 .

We shall now illustrate the use of the Table during testing. Before

starting the testing program, there are, of course, no failures and no successes and we start in the (0, 0) position. The lower entry is smaller than the upper, therefore a trial is conducted. If this trial is successful, move to the (0, 1) position (0 failures, 1 success); if the trial is a failure move to the (1, 0) position (1 failure, 0 successes). Either outcome, however, leads to a position within the dotted line so a second trial is conducted. The process continues until we are led to a position outside the dotted line. To get to a position outside the dotted line will require at least 2 trials and at most 9 trials. Hence, $N_m = 2$, and $N_M = 9$.

To confirm these numbers note that either 2 successive failures or successes will lead outside the line. Or, alternate success and failure will lead one diagonally down the Table to the (4, 4) position. This requires 8 trials. Once here, either another success or failure will lead one to a position which is outside the dotted line and not shown on the Table. Hence, the maximum number of trials is 9, and is the sum of the largest row and column designator plus one.

The testing program is thus completed when experimental results have led to crossing the dotted line of the Table as described above. Then the value of p_e is the number of successes divided by the number of trials. A comparison of p_e with the preassigned θ leads to the Terminal Decision d_1^t (convert) if $p_e > \theta$; and to the Terminal Decision d_2^t (develop), if $p_e < \theta$. If $p_e = \theta$, either decision can be made.

Table 3 (Simple W), is used in the same manner as Table 2. Note that the N_M is greater for the Simple W, specifically $N_M = 25$.

NUMBER OF SUCCESSES

NUMBER OF FAILURES	0	1	2	3	4	5	6	7	8	9	10	11	12
0	.33333 .04397	.11111 .03897	.03704 .02566	.01235 .01235	.00412 .00412	.00137 .00137	.00046 .00046	.00015 .00015	.00005 .00005	.00002 .00002	.00001 .00001	.00000 .00000	.00000 .00000
1	.11111 .03897	.25926 .05060	.11111 .04560	.04527 .03270	.01783 .01783	.00686 .00686	.00259 .00259	.00097 .00097	.00036 .00036	.00013 .00013	.00005 .00005	.00002 .00002	.00001 .00001
2	.03704 .02566	.11111 .04560	.20988 .05245	.10014 .04745	.04527 .03504	.01966 .01966	.00828 .00828	.00340 .00340	.00137 .00137	.00054 .00054	.00021 .00021	.00008 .00008	.00003 .00003
3	.01235 .01235	.04527 .03270	.10014 .04745	.17330 .05233	.08794 .04733	.04242 .03526	.01966 .01966	.00882 .00882	.00386 .00386	.00165 .00165	.00069 .00069	.00029 .00029	.00012 .00012
4	.00412 .00412	.01783 .01783	.04527 .03504	.08794 .04733	.14485 .05116	.07656 .04616	.03863 .03440	.01876 .01876	.00882 .00882	.00404 .00404	.00181 .00181	.00079 .00079	.00034 .00034
5	.00137 .00137	.00686 .00686	.01966 .01966	.04242 .03526	.07656 .04616	.12209 .04929	.06645 .04429	.03465 .03287	.01743 .01743	.00850 .00850	.00404 .00404	.00187 .00187	.00085 .00085
6	.00046 .00046	.00259 .00259	.00828 .00828	.01966 .01966	.03863 .03440	.06645 .04429	.10354 .04678	.05762 .04178	.03083 .03083	.01595 .01595	.00801 .00801	.00392 .00392	.00187 .00187
7	.00015 .00015	.00097 .00097	.00340 .00340	.00882 .00882	.01876 .01876	.03465 .03287	.05762 .04178	.08823 .04358	.04996 .03858	.02728 .02728	.01443 .01443	.00742 .00742	.00372 .00372
8	.00005 .00005	.00036 .00036	.00137 .00137	.00386 .00386	.00882 .00882	.01743 .01743	.03083 .03083	.04996 .03858	.07548 .04065	.04335 .03565	.02407 .02407	.01297 .01297	.00681 .00681
9	.00002 .00002	.00013 .00013	.00054 .00054	.00165 .00165	.00404 .00404	.00850 .00850	.01595 .01595	.02728 .02728	.04335 .03565	.06477 .03796	.03764 .03296	.02119 .02119	.01160 .01160
10	.00001 .00001	.00005 .00005	.00021 .00021	.00069 .00069	.00181 .00181	.00404 .00404	.00801 .00801	.01443 .01443	.02407 .02407	.03764 .03296	.05572 .03541	.03270 .03041	.01864 .01864
11	.00000 .00000	.00002 .00002	.00008 .00008	.00029 .00029	.00079 .00079	.00187 .00187	.00392 .00392	.00742 .00742	.01297 .01297	.02119 .02119	.03270 .03041	.04805 .03280	.02844 .02780
12	.00000 .00000	.00001 .00001	.00003 .00003	.00012 .00012	.00034 .00034	.00085 .00085	.00187 .00187	.00372 .00372	.00681 .00681	.01160 .01160	.01864 .01864	.02844 .02780	.04151 .02975

Solution Using Simple Weight Function - Table 3

CHAPTER V

VARIATION OF INPUT PARAMETERS

Tables similar to Table 2 and Table 3 may be prepared for other values of the basic parameters c and $\dot{\partial}$, for prescribed Weight Functions. From these tables N_M may be determined. Recall from page 38 that it is only the ratio C/W, which we called c, which is a parameter of the Tables.

The curves in Figures 17 and 18 show a plot of N_M for sets of values of the parameters as indicated. Figure 17 is for the Linear W and Figure 18 for the Simple W. These curves are plotted on semi-log paper. This was done since solutions are more easily obtained when c is varied by a factor of 10 thus the semi-log plot provides a more extended graph. The tables, from which the values of N_M were found, were obtained by means of a CRC model 102-A electronic digital computer, at the U. S Naval Postgraduate School. Such Tables may also be calculated with the aid of the mathematical tables in Pearson [6]. For details of such calculations see Tucker [1].

Note, from Figure 17 for any given value of ∂, as c decreases, the number of trials increases. This shows that if the cost of testing is decreased more trials are made before a decision is reached. Conversely, as the cost of each trial approaches the maximum value of W (maximum cost of wrong decision), fewer trials are conducted. Or, briefly, the more costly the testing, the fewer the number of trials.

Notice, further, that the number of trials required is very sensitive

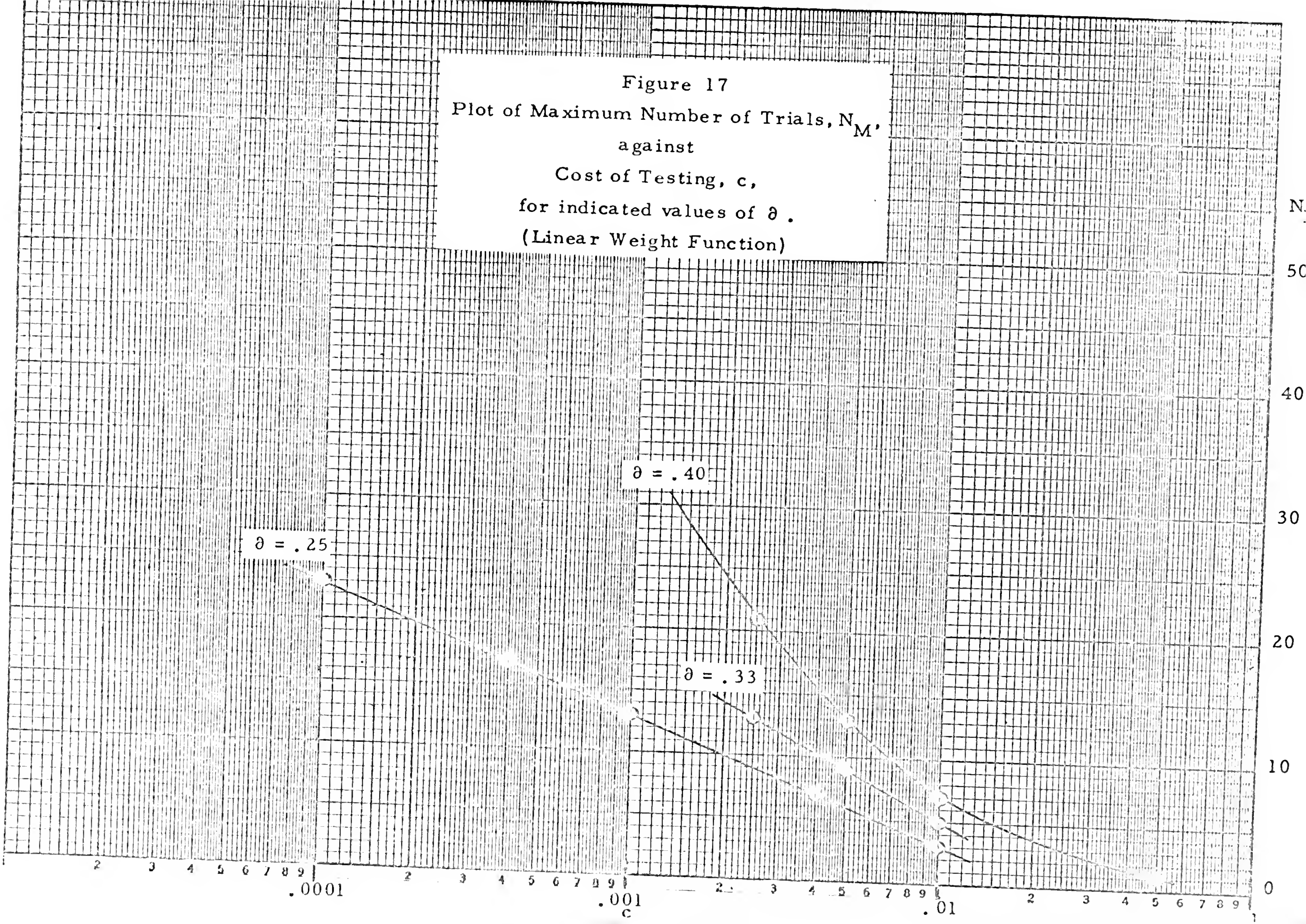


Figure 17
Plot of Maximum Number of Trials, N_M,
against
Cost of Testing, c,
for indicated values of ∂.
(Linear Weight Function)

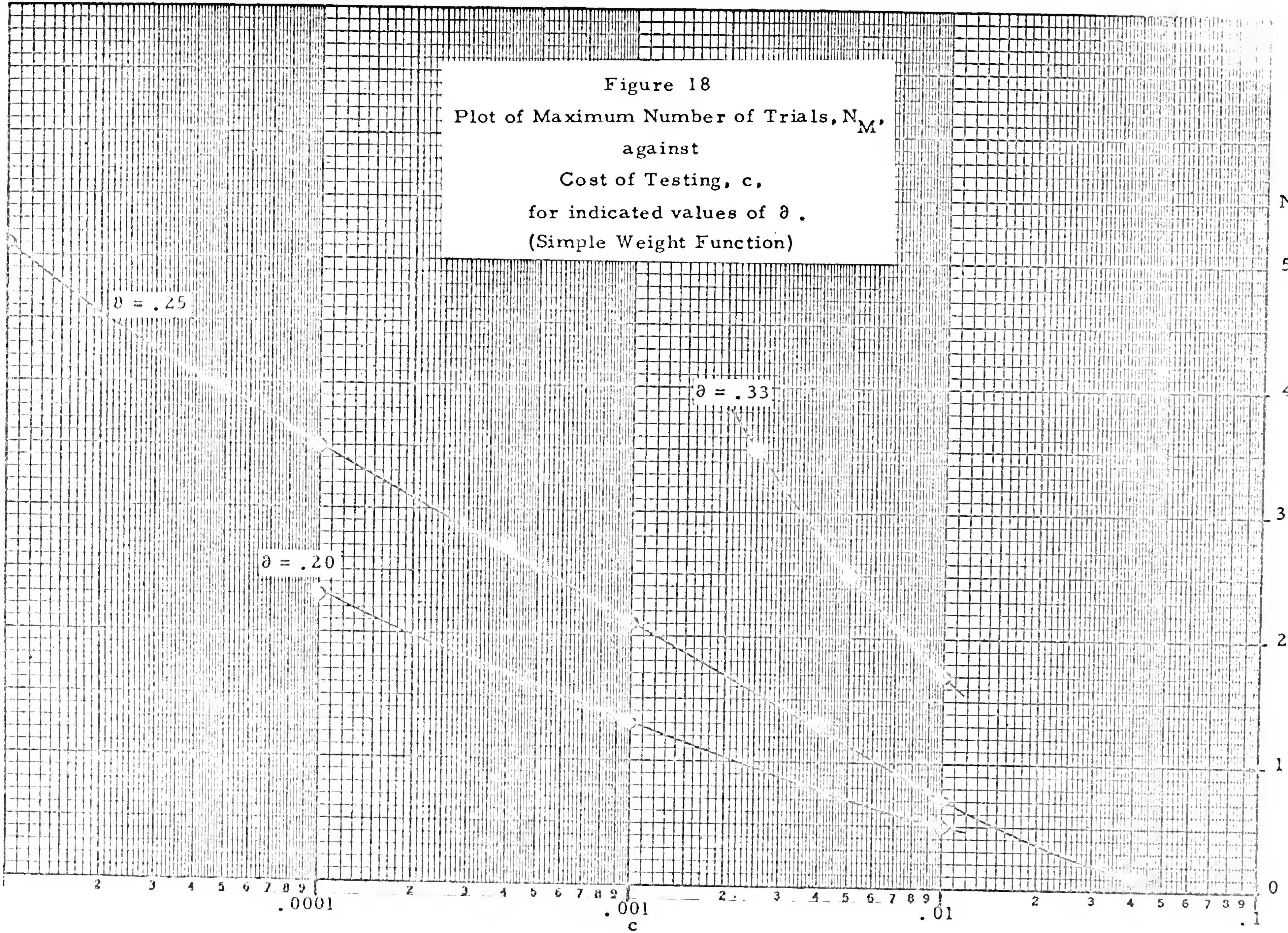


Figure 18
Plot of Maximum Number of Trials, N_M, against Cost of Testing, c, for indicated values of ∂.
(Simple Weight Function)

to changes in ∂ , particularly when ∂ is close to the value of θ . As ∂ approaches θ , the number of trials increases. A value of ∂ close to θ requires the theory to be more discriminating and this, in turn, requires more trials.

Figure 18 is similar to Figure 17 . The only difference being, that for given values of ∂ and c in Figure 17, the corresponding point in Figure 18 gives a larger value of n . In other words, for the same ∂ and c , the Simple W always results in more trials being required than indicated by the Linear W.

Summary

Statistical Decision Theory has been applied to a special problem, and two different ways in which the Weight Functions could be specified have been suggested. A value of N_M , the maximum number of trials required, is indicated by the solution. This number can be useful to naval planners as a basis for decision as to the number of missiles to be supplied to the testing agency for the tests. It should be noted that the value of N_M given by the solution will have to be increased to provide for those trials in which missiles are expended but, for some reason, the trial in invalidated.

It is assumed that naval planners contemplating the use of Decision Theory would have assistance from statisticians or mathematicians. Solutions can be extremely time consuming to produce manually, hence electronic computers will increase the practical usefulness of the theory. Dr. T. E. Oberbeck has obtained solutions for special cases using sim-

ple and linear Weight Functions, by programming the problem for the CRC model 102-A computer.[2]

For the use of naval planners, a more complete set of curves than those given in the chapter would facilitate the application of the theory. Such a project awaits further effort by research workers in this field.

In cases where the planner is unwilling to specify an a priori probability density function associated with $\dot{p}$ (see page 28), a Minimax solution of the problem may be obtained. For results on the average maximum number of trials see Breakwell [3].

[2]To be reported at the Fourth Annual Meeting of the Operations Research Society of America, in May 1956.

BIBLIOGRAPHY

1. Tucker, R. A. — INTRODUCTION TO STATISTICAL DECISION FUNCTIONS
U. S. Naval Postgraduate School, Monterey, Calif. 1955

2. Wald, A. — STATISTICAL DECISION FUNCTIONS
John Wiley and Sons, Inc. New York, 1954

3. Breakwell, J. V. — THE PROBLEM OF TESTING FOR THE FRACTION OF DEFECTIVES
Journal of The Operations Research Society of America, February 1954, Volume 2, Number 1

4. Mood, A. M. — INTRODUCTION TO THE THEORY OF STATISTICS
McGraw-Hill Book Co. Inc. New York, 1950

5. Statistical Research Group, Columbia University — SEQUENTIAL ANALYSIS OF STATISTICAL DATA: APPLICATIONS
Columbia University Press, New York 1945

6. Pearson, K. — TABLES OF THE INCOMPLETE BETA-FUNCTION
The University Press, Cambridge 1934

APPENDIX A

The graphs for sequential tests shown by Figures 6, 7 and 8 of Chapter III are constructed as follows.

(1) The ratio: $\lambda_R = \frac{p_{1n}}{p_{2n}} = \left(\frac{p_1}{p_2}\right)^s \left(\frac{q_1}{q_2}\right)^{n-s} = \left(\frac{p_1 q_2}{p_2 q_1}\right)^s \left(\frac{q_1}{q_2}\right)^n$

is calculated. Where n = number of trials
s = number of successes (hits)
$q = (1 - p)$

(2) The criteria defining the three regions are:

(a) If $\lambda_R \leqslant \frac{\beta}{1-\alpha}$, accept system II

(b) If $\lambda_R \geqslant \frac{1-\beta}{\alpha}$, reject system II

(c) If $\frac{\beta}{1-\alpha} \leqslant \lambda_R \leqslant \frac{1-\beta}{\alpha}$, continue testing

<u>For Case A</u>: $p_1 = .2$ $p_2 = .4$ $\alpha = .01$

$q_1 = .8$ $q_2 = .6$ $\beta = .1$

then the inequality (a) is

$$\left(\frac{3}{8}\right)^s \cdot \left(\frac{4}{3}\right)^n \leqslant \frac{\beta}{1-\alpha} \; ; \quad \frac{\beta}{1-\alpha} = \frac{1}{9.9}$$

Solving for s, by logarithms gives

$$s \geqslant 2.33 + .293n$$

Values of s and n which satisfy this inequality, define the acceptance region.

Similarly, inequality (b) is

$$\left(\frac{3}{8}\right)^{s}\left(\frac{4}{3}\right)^{n} \geqslant \frac{1-\beta}{\alpha} \quad ; \quad \frac{1-\beta}{\alpha} = 90$$

Solving for s gives

$$s \lesssim -4.58 + .293n$$

Values of s and n which satisfy this inequality, define the rejection region .

(3) The parallel lines in Figure 6 are obtained by plotting the straight lines

$$s = 2.33 + .293n$$

$$s = -4.58 + .293n$$

In the same manner, the straight lines in Figure 7 and 8 are obtained.

<u>For Case B</u> :

$$s = 3.05 + .293n$$

$$s = -4.64 + .293n$$

<u>For Case C</u> :

$$s = 1.28 + .387n$$

$$s = -2.51 + .387n$$

Thesis
S5723 Smith
c.2
Application of statistical methods to naval operational testing.